Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples

Bista, Iliana; Carvalho, Gary; Tang, Min; Walsh, Kerry; Zhou, Xin; Hajibabaei, Mehrdad; Shokralla, Shadi; Seymour, Mathew; Bradley, David ; Liu, Shanlin; Christmas, Martin; Creer, Simon

Molecular Ecology Resources

DOI:
10.1111/1755-0998.12888

Published: 01/09/2018

Peer reviewed version

MISS ILIANA BISTA (Orcid ID : 0000-0002-6155-3093)

Article type : Resource Article

Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples

Authors

Iliana Bista[1, 2*], Gary R. Carvalho[2], Min Tang[3], Kerry Walsh[4], Xin Zhou[3], Mehrdad Hajibabaei[5], Shadi Shokralla[5], Mathew Seymour[2], David Bradley[6], Shanlin Liu[7, 8], Martin Christmas[4], Simon Creer[2]

1: Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

2: Bangor University, School of Biological Sciences, Molecular Ecology and Fisheries Genetics Laboratory, LL57 2UW

3: Department of Entomology, China Agricultural University, Beijing 100193, People's Republic of China

4: Environment Agency, Horizon House, Deanery Road, Bristol BS1 5AH, UK

5: Department of Integrative Biology & Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1

6: APEM LTD, Heaton Mersey, Stockport, Cheshire, SK4 3GN

7: Natural History Museum of Denmark, Øster Voldgade 5-7, 1350 København K

8: BGI-Shenzhen, Shenzhen, 518083, China

Keywords: metabarcoding, metagenomics, biomass, genome-skimming, invertebrates, biodiversity

Running title: Shotgun vs. amplicon invertebrate communities

Corresponding author*

*Iliana Bista, ilianabista@gmail.com

Simon Creer, s.creer@bangor.ac.uk

Abstract

New applications of DNA and RNA sequencing are expanding the field of biodiversity discovery and ecological monitoring, yet questions remain regarding precision and efficiency. Due to primer bias, the ability of metabarcoding to accurately depict biomass of different taxa from bulk communities remains unclear, while PCR-free whole mitochondrial genome (mitogenome) sequencing may provide a more reliable alternative. Here we used a set of documented mock communities comprising 13 species of freshwater macroinvertebrates of estimated individual biomass, to compare the detection efficiency of COI metabarcoding (3 different amplicons) and shotgun mitogenome sequencing. Additionally, we used individual COI barcoding and *de novo* mitochondrial genome sequencing, to provide reference sequences for OTU assignment and metagenome mapping (mitogenome-skimming) respectively. We found that even though both methods occasionally failed to recover very low abundance species, metabarcoding was less consistent, by failing to recover some species with higher abundances, probably due to primer bias. Shotgun sequencing results provided highly significant correlations between

read number and biomass in all but one species. Conversely, the read-biomass relationships obtained from metabarcoding varied across amplicons. Specifically, we found significant relationships for 8 out of 13 (amplicons B1FR-450bp, FF130R-130bp) or 4 out of 13 (amplicon FFFR, 658bp) species. Combining the results of all three COI amplicons (multi-amplicon approach) improved the read-biomass correlations for some of the species. Overall, mitogenomic sequencing yielded more informative predictions of biomass content from bulk macroinvertebrate communities than metabarcoding. However, for large scale ecological studies, metabarcoding currently remains the most commonly used approach for diversity assessment.

Introduction

The accurate qualitative and quantitative assessment of biodiversity is essential in order to understand biodiversity and ecosystem function relationships, especially in the face of rapid biodiversity loss (Loreau & de Mazancourt, 2013). However, evaluating the speed and scale of ecosystem degradation is limited by the use of traditional taxonomic approaches that typically require high levels of expertise and are labour-intensive. Moreover, accurate species identification is frequently not possible in cases of damaged or immature specimens (Jackson et al., 2014; Sweeney, Battle, Jackson, & Dapkey, 2011). In biomonitoring, the robust quantification of community composition enables detection of both spatial and temporal variations in the biological community, and by extension, the wider ecosystem (Cranston, 1990). To expedite biomonitoring frameworks, international directives, such as the EU Water Framework Directive (WFD) have been established. Initiatives such as the WFD require ecological status classification of water bodies which are

underpinned by the taxonomic identification of organisms from routine monitoring (Collins, Ohandja, Hoare, & Voulvoulis, 2012).

The advent of high throughput sequencing technologies (HTS) is revolutionising biomonitoring by increasing the throughput and taxonomic information that can be recovered (Baird & Hajibabaei, 2012). The most commonly used taxonomic groups for such studies include various invertebrate taxa, such as benthic macroinvertebrates for freshwater ecosystems (e.g. Gibson et al., 2014; Gibson et al., 2015; Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011; Pfrender et al., 2010; Shokralla et al., 2015). Similarly, terrestrial invertebrate taxa have been used, from soil or leaf litter (Yang et al., 2014), or from above ground invertebrate sampling (Malaise traps) (Ji et al., 2013). More recent work encompasses the detection of biodiversity from aqueous environmental DNA (eDNA) (Bista et al., 2017; Ficetola, Miaud, Pompanon, & Taberlet, 2008; Mächler, Deiner, Steinmann, & Altermatt, 2014; Seymour et al., 2018).

Most studies using HTS for biodiversity assessment of bulk samples to date utilise PCR amplicon sequencing of one or more marker genes. Amplicon sequencing is often referred to as metabarcoding and is commonly applied for the analysis of bulk/environmental community samples (Creer et al., 2016; Yu et al., 2012). Commonly used markers include the mitochondrial Cytochrome C Oxidase Subunit I (COI) barcode region (Hebert, Ratnasingham, & de Waard, 2003) and ribosomal RNA regions such as 16S (Epp et al., 2012) for animals, or RbcL and ITS2 for plants (Fahner, Shokralla, Baird, & Hajibabaei, 2016). Due to intermediate PCR steps, it is considered that metabarcoding produces bias in relation to accurate taxonomic representation of diversity in bulk samples (Hajibabaei, Spall, Shokralla, & van Konynenburg, 2012; Yu et al., 2012). In fact, it has been reported that PCR bias might alter

the species biomass ratio or produce inaccurate representation of relative abundance of species (Piñol, Mir, Gomez-Polo, & Agustí, 2015), where primer-template mismatches might introduce mis-representation of particular groups (Clarke, Soubrier, Weyrich, & Cooper, 2014; Elbrecht & Leese, 2015). On the other hand, correlations between amplicon read number and biomass have also been reported (Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017; Hiiesalu et al., 2014; Kelly et al., 2014). Optimisation of metabarcoding pipelines, use of multiple primer pairs, and combination of multiple amplicons from the same region has been suggested to improve species richness recovery and biomass estimations (Gibson et al., 2014; Hajibabaei et al., 2012; Zhan, Bailey, Heath, & Macisaac, 2014). Such improvements to methodological aspects are critical to standardise where possible, not only to improve comparability of independent data sets, but importantly, also to promote accessibility and eventual uptake of such technologies.

The more recently introduced mitochondrial metagenomics (mitogenomics) uses high throughput Illumina sequencing of whole mitochondrial genomes from bulk samples (Crampton-Platt, Yu, Zhou, & Vogler, 2016). Some current applications involve characterisation of bulk samples for ecological assessment (Tang et al., 2015), and phylogenetic reconstruction of multiple species simultaneously (Gillett et al., 2014). This approach has been suggested as an alternative to PCR-based metabarcoding (Zhou et al., 2013), advocating that the absence of initial PCR amplification step will reduce PCR related bias. Specifically, it is suggested that mitogenomic sequencing results in more accurate biomass to reads relationships and possibly more reliable representation of species biomass in bulk samples. The efficiency of this approach could be related to the assembly method

used, and whether reference sequences are available for mapping (Gómez-Rodríguez, Crampton-Platt, Timmermans, Baselga, & Vogler, 2015).

Here, we provide a direct comparison between metabarcoding and mitogenomics, and investigate the ability of each method to depict the richness and biomass presence of macroinvertebrate species in bulk extracted samples. Ten artificial communities were used, comprising 13 common macroinvertebrate species, for which all specimens were individually measured for their biomass. The samples were selected to represent a typical freshwater biomonitoring sample collected for environmental assessment purposes, including various taxonomic orders, while accounting for availability of large numbers of individuals per target species. Additionally, the replicate communities were designed to comprise gradients of biomass, from as low as a single specimen up to multiple specimens depending on the species (e.g. step increase of specimens per community from 1 to 33 for *Potamopyrgus antipodarum*, and 1 to 17 specimens for *Radix balthica*), to represent a range of natural densities. DNA was extracted from pooled whole bodies of invertebrates, unlike previous work (e.g. Crampton-Platt et al., (2015); Gómez-Rodríguez et al., (2015)). The community DNA extracts were amplicon sequenced for three fragments of the COI gene on Illumina MiSeq, and shotgun sequenced on Illumina HiSeq (see Figure 1 for overview of experimental workflow). Possible limitations and advantages of each method were considered, to evaluate their performance. Furthermore, the overall applicability of each method was assessed while providing suggestions for future improvements. Ultimately, we aim to provide a comprehensive evaluation of the two methods and troubleshoot their future utility for ecological applications in biodiversity and freshwater ecosystem monitoring.

Materials and Methods

Specimen collection and morphological measurements

Specimens for this work were collected from areas of Somerset and Suffolk, UK (September – October 2014), and were identified to species level (by county surveyors and APEM Ltd, U.K.). Specimens were preserved in absolute ethanol, and stored in a dark and cool environment for up to 6 months until morphological measurements and DNA extraction. In total, 13 species were used for analysis, including 8 species of Gastropoda and one of each from: Hemiptera, Isopoda, Amphipoda, Ephemeroptera and Coleoptera (Table S1, Figure S1), which were selected to represent a typical biomonitoring sample.

Body measurements were performed based on published work using the following methodologies: callipers for larger species (*Notonecta glauca, Asellus aquaticus, Gyrinus marinus, Ephemera danica*) and a microscope fitted with an ocular micrometre for smaller species (e.g. *Potamopyrgus antipodarum*). Software Image Pro (Media Cybernetics, Rockville, USA) was used for the amphipod *Gammarus pulex*, to facilitate accuracy by accounting for the curvature of specimens. Published regressions based on length to mass measurements were used for estimation of biomass for each species (Table S2), taking into account geographic region and ecosystem type wherever possible, as these parameters could produce intraspecific variations in development rates (Mährlein, Pätzig, Brauns, & Dolman, 2016). Conversion of length to mass is considered superior to other methodologies, such as determination of biovolume or weighing of specimens, due to increased precision and speed (Benke, Huryn, Smock, & Wallace, 1999). Additionally, direct weighing of preserved specimens might be inaccurate due to loss of dry mass during preservation (Benke et al., 1999).

DNA barcode Reference Library

Individual specimens were extracted and sequenced for the COI barcode region using universal metazoan primers (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994). Extraction was performed with DNeasy Blood & Tissue kit (QIAGEN) (arthropods), and CTAB chloroform extraction protocol (gastropods), due to best suitability of protocols for different tissue types. High quality barcodes were obtained from all species (Table S1), except *E. danica* (Ephemeroptera), for which barcode sequencing was not successful (for similar findings regarding Ephemeroptera barcoding see Morinière et al. 2017). Sanger generated sequences were edited using CodonCode Aligner v.3.7.1 (CodonCode Corporation, Massachusetts) and aligned with ClustalW in MEGA v5 (Tamura, Dudley, Nei, & Kumar, 2007).

Design of mock communities

Ten communities were created containing either 13 or 14 species (including control species, see below), except for two species, where low numbers allowed representation in only 6 (*N. glauca*) or 9 (*P. fontinalis*) of the communities. Specimens ranged from 139 to 157 individuals, each at a gradient of biomass presence from a single to multiple specimens (Table 1, Table 2). The total sum of specimens across 10 communities was 1479. Two positive controls were used to assess the quality of sequencing performance across mock communities. First, three specimens of *Drosophila melanogaster* were added in each community, to assess extraction efficiency. Second, DNA of the butterfly *Mycalesis mineus*, which was separately extracted (previously assembled by Tang *et al.* 2014), was added in each community extract at 1% concentration (at BGI-Shenzhen, China), to account for

variability in DNA subsampling before shotgun sequencing. The two control species were selected due to existing reference mitogenome information.

DNA extraction for reference mitogenomes and mock communities

For the construction of individual shotgun reference genomes for each species, high quality genomic DNA was extracted from leg or muscle tissue of a single specimen using the Qiagen DNeasy Blood and Tissue extraction kit [final elution in 50µl PCR Grade water (Roche)] and DNA concentration and quality was assessed with dsQubit DNA assays and agarose gel electrophoresis. A minimum amount of 2.5µg total DNA was used for shotgun sequencing.

For the mock communities, DNA was extracted from the entire bodies of invertebrates, in bulk. First, ethanol preservative was removed and then specimens were allowed to dry at 37^{o}C for 2 hours. The dried specimens were ground using sterile mortar and pestle sets, then transferred into 50ml Power Bead tubes from the Power Max Soil DNA Isolation Kit (MO-BIO) and vortexed at high speed for 5min. Lysis was performed by incubation for 3h at 65°C in a shaker at medium speed, with the addition of 450µl of Proteinase K (20mg/ml) (Sigma-Aldrich). Subsequent steps followed manufacturer's instructions. For final elution (total volume 4ml), the columns were incubated for 30min at room temperature and centrifuged at 2500g for 5min. The elution step was repeated a second time to maximise yield (Figure S2, Table S3). For the purposes of this work, we did not use any mtDNA enrichment method in order to avoid skewing the biomass proportions of species between samples.

Metabarcoding - Primer selection and amplicon library preparation

Metabarcoding was performed through sequencing of three amplicons of the COI barcode region using the following primer pairs: 1.) amplicon FFFR using the Folmer primers (Folmer et al., 1994) (658bp); 2.) amplicon FF130R using Folmer forward primer – I-130R primer (130bp); and 3.) amplicon B1FR using I-B1 forward primer and Folmer reverse (450bp) (I-B1 primer was modified from Hajibabaei *et al.* (2012), and I-130R was modified from Meusnier et al. (2008) both degenerate, modified for use with macroinvertebrate communities] (Figure 2 and Table 3). These primers were selected based on visual comparison with aligned barcode sequences (Appendix I, sequence alignment of COI barcodes and primers I-B1 and I-130R), and amplification performance tests on mock community extracts. Libraries were prepared using a two-step PCR protocol with final indexing step, to minimise the effects of variant index sequences on the amplification efficiency of each community (Berry, Mahfoudh, Wagner, & Loy, 2011; O'donnell, Kelly, Lowell, & Port, 2016). Round 1: amplification was performed using only the target specific COI primer, Round 2: purified PCR product from round 1 was used as template using the target specific primers with added Illumina tails, and a final step was used for indexing of all amplicons from round two with Illumina indexes. The samples were sequenced on an Illumina MiSeq using MiSeq V2 reagents (500 cycles) following the (2x250) paired-end protocol (PE), at the Biodiversity Institute of Ontario, University of Guelph (Canada). Detailed PCR protocols are available in Appendix II. Throughout the library preparation, appropriate laboratory precautions were taken to establish clean conditions for the analysis of community DNA. Protocols included sterilization of equipment by autoclaving or bleaching, and PCR set ups were performed in a dedicated clean cabinet.

Amplicon data analysis

Quality control and analysis of the Illumina MiSeq sequences was performed according to Bista *et al.* (2017), using a USEARCH v7 (Edgar, 2010) custom pipeline for filtering, sorting, de-replication and merging of sequences. For amplicons 2 (FF130R, 130bp) and 3 (B1FR, 450bp), the forward and reverse reads were merged with a 25bp minimum overlap. For amplicon 1 (FFFR, whole barcode region 658bp) only the forward reads were used (truncation at 230bp), based on FastQC (www.bioinformatics.babraham.ac.uk) results (phred score >25). The truncation strategy was selected because the length of the original amplicon (658bp) did not allow sufficient overlap between the forward and reverse reads due to the current limitations of Illumina 2x250bp MiSeq chemistry. Chimeras were removed with the *de novo* option in USEARCH, and a 97% similarity level was used for OTU clustering and generation of an OTU table in USEARCH (the dominant OTU centroids in this case are selected through the most abundant sequences). The OTU level of similarity was used as an approximate average value for characterisation of the diverse taxa present in the community samples. Similarly, taxonomic assignment was performed as in Bista *et al.* (2017), using Quantitative Insights In Microbial Ecology (QIIME) (Caporaso et al., 2010), and BLAST+ (megablast) (Camacho et al., 2009) with ≥98% similarity cut-off, against the downloaded NCBI COI barcodes, and our custom generated COI barcodes. All analyses including USEARCH, QIIME and BLAST were performed using High Performance Computing (HPC) Wales systems. The identified OTUs were further checked for stop codons and insertions in MEGA6 (Tamura et al., 2007), and through phylogenetic analysis using a Neighbour-Joining (NJ) method (Saitou & Nei, 1987). Metabarcoding library preparation and data analysis was performed at Bangor (UK), and MiSeq sequencing was run at BIO Guelph (Canada).

Construction of reference mitogenomes and mitogenome skimming

Genomic DNA extracted from individual specimens was used for generation of reference mitogenomes. For each species, a library with insert size of 200bp was constructed following manufacturer's instructions (Illumina, Nextera). The 12 individual species libraries were pooled and 100bp PE sequenced on a whole lane of Illumina HiSeq2000. Library construction and assembly of reference mitogenomes as well as mitogenome skimming and analysis of mock community samples were performed at BGI, Shenzhen, China. For reference mitogenomes raw data from each species were filtered as previously described in Zhou et al. (2013), Tang et al. (2014) and Tang et al. (2015), removing reads with low quality or adaptor contamination. Clean data were assembled into scaffolds using SOAPdenovo-Trans (-K 71) (Xie et al., 2014) and IDBA-UD (Peng, Leung, Yiu, & Chin, 2012). Assembled sequences were annotated following Tang et al. (2015), to identify candidate mitogenome sequences, which were used for mitogenome reference construction. Subsequently manual correction and checking were carried out as described by Tang et al. (2014). Thirteen protein-coding genes (PCG) were extracted from all mitogenomes, and each of them were aligned with corresponding reference protein-coding genes from 4 arthropod species (*Macrogyrus oblongus, Gammarus duebeni, Ligia oceanica* and *Siphlonurus immanis*) and 3 mollusc species (*Biomphalaria tenagophila, Physella acuta* and *Oncomelania hupensis*) using CLUSTALW 2.1 (Thompson, Higgins, & Gibson, 1994). The translation frame was checked in MEGA6 (Tamura et al., 2007) to correct gap length generated inside protein-coding genes by the assembly program when constructing scaffolds based on PE reads. In addition, the original read-mapping was done and monitored using BWA 0.6.2 (Li & Durbin, 2009) and SAMTOOLS 0.1.19 (Li et al., 2009) respectively following Tang *et al.* (2014, 2015).

Each of the 10 community DNA samples was used for construction of 200bp insert-size libraries, which were 100bp PE sequenced on two lanes of a HiSeq2000 (ca. 2-3 Gb per sample), according to manufacturer's instructions. Filtered data were aligned onto the 12 previously constructed reference mitogenomes by BWA, and reads that uniquely mapped onto the references with 100% read coverage and ≥99% identity were considered as reads from the focal species.

Statistical and community analysis

To allow statistical comparisons between samples, we performed normalisation of sequencing data per mock community, while accounting for variation of sequencing depth. For the shotgun data, normalisation was performed by mitogenome length and mito-ratio (MitoNorm), and as proportion of mitochondrial reads on total reads per mock community (pShotgun) (see detailed information in Tang et al., (2015)). For the amplicon data normalisation was performed as amount of OTU reads from the total number of reads per mock community, for each amplicon (target_species_reads/total_community_reads). To select the best model explaining the relationship between number of reads and biomass (mg), linear and logistic models were compared for each species and sequencing methods. The best model was selected using Akaike information criterion (AIC) (Hu, 1987). All statistical analyses, including calculation of model parameters, were performed using the program R Core Team (2015).

To visualise community variation across sequencing treatments for amplicon data, nonmetric multidimensional scaling was performed (nMDS), using the metaMDS function in the *vegan* package in R (version 3.3.0), based on calculation of the Bray-Curtis dissimilarity index as a measure of relative abundance, based on species level data. Additionally, the

function “ordispider” in package vegan was used to connect the same communities (resulting from different sequencing treatments) on the ordination plot. The software PRIMER-E v6 (K. R. Clarke & Gorley, 2006) was also used to examine differences in community composition between sequencing methods (nMDS, Bray-Curtis).

Results

Sequencing results

For metabarcoding data, the total number of sequencing reads obtained after quality control was 1,430,531, sequenced on a fraction of an Illumina MiSeq lane. Each amplicon produced the following total number of reads (Mean ± SD), 1.) FFFR1: 248,776 (24,878 ± 16,815), 2.) FF130R: 1,004,530 (100,453 ± 87,366), 3.) B1FR: 177,225 (17,722.5 ± 24,418) (Table S4). After OTU clustering, for each amplicon, we obtained 49 (FF130R), 20 (FFFR) and 14 (B1FR) OTUs respectively. Collapsing of multiple OTUs per species was used to account for intraspecific diversity in our data although the observed intraspecific diversity amongst OTUs assigned to the same species was generally low (Table S5).

We sequenced the reference mitochondrial genomes for 12 out of 13 species used in this experiment; species *A. vortex* was not included in the run due to low quality of extracted DNA. Total length of mitochondrial DNA assemblies ranged between 13,326 - 16,159 bp, with three species also achieving circular genomes (*Notonecta glauca, Physa phontinalis and Gyrinus marinus*) (Table S6). The average mitochondrial genome length was 14,760 bp.

The amount of data attributed to mitochondrial reads compared to the total reads per species (mito-ratio) varied largely between species, ranging between 0.011% (*G. pulex*) and 0.692% (*E. danica*), with an average mito-ratio of 0.19%. The average sequencing depth per

mitogenome was 177.47X (min depth: 6.4X – *G. pulex*, max depth: 670.4X – *E. danica*) (Table S6). Shotgun sequencing of the mock community samples returned 23,984,200,200 total number of reads with an average of 2,398,420,020 reads (Figure S3, Table S4).

For the two positive controls, *D. melanogaster* returned an average of 344.2 (±51.3) reads, and for *M. mineus* an average of 787.3 (±125.2) (Figure S4) (read numbers for mitochondrial genomes only). The latter was significantly correlated with the number of reads achieved per sample (R^2 = 0.717, p = 0.002), while no significant relationship was found for the *D. melanogaster* read number vs. total number. For amplicon sequencing, only the *D. melanogaster* positive control was used. Significant relationships between the positive control sample and the total number of reads were found for two of the amplicons (B1FR: R^2 = 0.939, p = 0) (FF130R: R^2 = 0.610, p = 0.008), but not for the whole COI region amplicon (FFFR1).

Detection rates per species

Cases of false negatives and false positives were found for both sequencing methods, based on a lower cut-off of reads present (defined as zero for indicating presence/absence of species). The proportion presence of false negatives is reported here based on number of expected (known) occurrences (presence in sample) for each species in the communities. Occurrences were calculated normally as 10 per species (10 communities), except for species *P. fontinalis* (9 occurrences) and *N. glauca* (6 occurrences) [(11sp. x 10) + (1sp. x 6) + (1sp. x 9) = 125 total occurrences/cases] (based on counts from Table 1). For this step, we were considering the number of species across all 10 communities to increase statistical power of calculations.

The shotgun approach failed to detect the presence of species in the community samples in 7 out of 125 cases (5.6%), for 5 species. Generally, the false negatives with this method occurred only for the lowest and second lowest amount of biomass present for the rare species. For the three amplicons, false negatives occurred in a.) 3 cases (2.2% in 3 species) for FFFR, b.) 6 cases (4.5% in 3 species) for FF130R, and c.) 7 cases (5.6% in 5 species) for B1FR. (Table 1, Table S7). Here false negatives appeared not only for the lowest biomass of species, but also when up to 10 (FF130R, FFFR), 13 (FFFR) or 17 (B1FR) specimens were known to be present in that community. Overall, false negatives mostly came from gastropod species except, *G. pulex* (amphipod, 2 cases) and *E. danica* (mayfly, 1 case). False positives were detected for species *N. glauca* in two cases (FFFR amplicon), with 111 and 34,511 reads detected in communities 7 and 9 respectively, and for species *P. fontinalis* with 1,204 reads in community 10 (FFFR amplicon). The presence of false positives here could be attributed to cross-contamination due to carry over from species during common storage of specimens.

Associations between biomass and number of reads

Statistical model investigations suggested that both logistic and linear models were appropriate for characterising the number of reads to biomass relationships, with the model type generally linked to species across the different sequencing methods (Table 4). The relationship of reads with biomass was examined individually for each sequencing treatment (three COI amplicons, sum of all amplicon data, and shotgun data) and each species, and plotted with the appropriate best-fit model (Figure 3, Figure 4, Figures S5-S8). Additionally, the total data obtained for each community were simultaneously plotted

against specimen biomass to visualise overall trends per community (Shotgun and sum of amplicon data, Figures S9-S10).

Positive and mostly significant relationships were found between biomass – shotgun reads (11 out of 12 species with 1 trending towards significance; $p = 0.08$). Metabarcoding results varied across amplicons, but we found significant positive read – biomass relationships for 4 - 8 out of 13 species (Table 4). All species presented positive relationships, with the exception of *E. danica* which presented negative reads -biomass relationship, but only for the FFFR amplicon generated using the universal Folmer primers (Figure S7). We also investigated the relationships obtained when adding up the reads obtained from all three amplicons vs. known biomass, as sum of amplicon data (Table 4, Figure 4). In this case, the detection rates improved, by removing false negatives. Nevertheless, significant reads – biomass relationships were found in only 8 species, which was an improvement in comparison to using only FFFR (Folmer) amplicon data, but not in comparison to the other two amplicons (Table 4).

Community analysis

Comparison between the three COI amplicons on the nMDS showed grouping of the same communities along the vertical axis with the exception of communities 9 and 10 (Figure S11). Such findings suggest a qualitatively similar community composition in the results obtained by the different amplicons. Simultaneous plotting of amplicon and shotgun data (Figure S12a) shows each sequencing treatment separated along the horizontal axis. When the amplicon data were plotted as a sum (SumAmplicon) (Figure S12b) against the shotgun reads we could again only observe vertical separation of the groups, although in this case, much clearer than when the individual amplicons were plotted. Moreover, the

similarity ranking of communities was almost identical for the two types of sequencing (see order of communities as B, C, G, I etc.).

Discussion

Here we used two HTS based methods for DNA based biodiversity analysis, metabarcoding and shotgun mitogenomics, to characterize species composition and biomass of bulk invertebrate samples. Unlike previous work (Crampton-Platt et al., 2015; Gómez-Rodríguez et al., 2015; Yu et al., 2012) we extracted whole bodies in bulk and also used exact biomass measurements for each specimen used, through a structured design of replicates which allowed testing of different biomass scenarios (from single specimen to multiple per species). Detection of rare taxa proved challenging for both methods, as they failed to detect low biomass species in several cases, while metabarcoding also misrepresented higher biomass species as well. Our results suggest that using shotgun mitogenomic sequencing provides a more consistent and representative estimate of the relationship between reads and biomass from bulk macroinvertebrate samples, compared to amplicon metabarcoding of the COI gene. When considering the single amplicon approach, the metabarcoding data did not provide accurate quantitative information on the biomass of a large proportion of the species in bulk macroinvertebrate samples, although the accuracy of the method slightly improved when results from all three amplicons were combined.

Sequencing performance and sample coverage

Sufficient coverage of reference mitogenome sequencing was achieved (Table S6), enabling assembly of almost complete reference mitogenomes. Previous reports suggest that 10X coverage would allow shotgun mitogenome assembly (Zhou et al. 2013), but even

at 6.4X coverage for *G. pulex*, we assembled 13,326 bp, at an estimated 83-85% of expected length, by comparison to other available closely related species [see species *Gammarus duebeni*, 15,651 bp, in Krebes & Bastrop (2012) and *Gammarus roeselli* 15,989 in Macher, Zizka, Weigand, & Leese (2017)]. The low coverage achieved for *G. pulex* could be possibly related to large nuclear genome for this species (as reported for related species with nuclear genome size 8.5-10.5 pg (Jeffery & Gregory, 2014), which accounted for the largest proportion of sequencing reads (mito-ratio 0.011%, Table S6). By using existing barcode sequences as baits for mapping, our pipeline allowed lower sequencing coverage to be sufficient compared to *de novo* assembly (read based approach) (Crampton-Platt et al., 2016).

For the metabarcoding work, sequencing coverage varied among amplicons, with the shorter amplicon (FF130R, 130bp) obtaining higher number of reads compared to the other two amplicons (B1FR, 450bp; FFFR, 658bp) (Table S4). Such variation in the depth of sequencing could be attributed to Illumina MiSeq sequencing preferentially amplifying shorter reads when sequencing mixed length amplicons or variable efficiency of primer binding (Aird et al., 2011). Additionally, this variation could be attributed to different amplification efficiency when different primer pairs are used across a range of taxa. Normalising library contents during sequencing (according to size of molecules included) should therefore be taken into consideration when multiple amplicons are sequenced in the same run.

Sequence reads – biomass relationships for both methods

The majority of species presented positive relationships of biomass with the read data for both methods, while only one species showed negative relationship (species *E. danica,* in

metabarcoding data). The *E. danica* reverse trend was found for the FFFR (658bp) amplicon (Figure S7), which was sequenced using the universal Folmer primers (Folmer et al., 1994). The same species also failed to amplify during individual barcoding (Table S1), suggesting that the results are likely to be related to primer incompatibility. In some cases, the use of a logistic model was a better descriptor of the relationship between biomass and read number compared to linear models (Table 4). In most cases the number of reads per amplicon increased logistically with increasing biomass, suggesting a biological link between amplicon read number and species biomass. Both linear and logistic models have been used for the representation of reads to biomass relationships in published metabarcoding and mitogenomic studies (Doi et al., 2017; Elbrecht & Leese, 2015; Lacoursière-Roussel, Rosabal, & Bernatchez, 2016; Tang et al., 2015; Zhou et al., 2013).

Perspective on mitogenomic based analysis of biodiversity

We performed mitogenome skimming based on custom generated reference mitochondrial genomes, which allowed use of reduced sequencing depth. In the absence of a reference genome, species *A. vortex* was not included in downstream analysis of shotgun data. Other studies using mitogenomic sequencing to characterise assemblages of leaf beetles (Gómez-Rodríguez et al., 2015) or mass-trapped arthropods (Choo, Crampton-Platt, & Vogler, 2017), have also shown that mitogenome assembly based on reference sequenced genomes outperforms the *de novo* approach (without reference library) in accuracy and recovery of diversity. Additionally, availability of reference mitogenomes during analysis allows easier detection and removal of Nuclear Mitochondrial pseudogenes (NUMTs) (Bensasson, Zhang, Hartl, & Hewitt, 2001) from shotgun sequencing data (Tang et al., 2014).

For the core analyses, the shotgun reads were normalised according to proportion of reads, and mito-ratio, which accounted for the variability of mitochondrial sequencing effort compared to the total amount of sequencing reads per community. Our investigation of normalisation methods showed similar findings between reads normalised according to mito-ratio and proportion of total reads (Table 4). Normalisation of sequencing data is used to account for different DNA concentrations produced by the variability of body size (Gillett et al., 2014), number of individuals and relative abundance in communities. A variety of normalisation options are available (Weiss et al., 2017). Similar normalisation strategies as used in the present work have also been used in other studies (e.g. in Tang *et al.* (2015)).

Mitogenomic sequencing currently uses a very small fraction of the total sequencing data, as the genomic DNA represents the largest amount of total DNA in the sample (Zhou et al., 2013). Depending on the taxon, the genomic to mitochondrial DNA ratio (mito-ratio) might vary, but generally approximately 99% of the reads are attributed to genomic DNA, leaving only 0.5-1% of the data to be used (for insects the mito-ratio is 0.5% or lower) (Zhou et al., 2013). Attempts to generalise the expected genomic to mitochondrial DNA ratio are difficult as further work on a wider variety of taxa is necessary (Crampton-Platt et al., 2016). In the present work mito-nuclear ratio ranged between 0.011% (*G. pulex*) and 0.692% (*E. danica*), with an average of 0.19% Table S6), which was lower than the previously described, nevertheless it did not seem to present a huge challenge for the method, as was demonstrated herein.

To enhance the mtDNA contribution to the data, enrichment via centrifugation (during DNA extraction) (Macher et al., 2017; Zhou et al., 2013), and oligonucleotide capture array (Liu et al., 2016) have been used. In Zhou et al. (2013), only a moderate increase of mtDNA reads was achieved, which accounted for about 0.5% of the total data, while Macher et al.

(2017) report a 129 to 140-fold enrichment of mtDNA through centrifugation. Use of capture arrays, designed based on 379 mitochondrial genomes (Liu et al., 2016) increased the mitochondrial ratio by 100-fold compared to previous attempts (mtDNA reads accounted for ca. 42% of the sequencing data after enrichment). Additionally, use of a capture array maintained the original ratio of species biomass in the sample, with a few variations depending on the phylogenetic distance of the test sample species composition compared to the species used for designing the array. The accuracy of capture arrays could be limited by the availability of sequencing information for the target organisms used for designing the probes (Hajibabaei, Singer, Clare, & Hebert, 2007) or due to occasionally picking-up non target taxa (Liu et al., 2016).

Applications of mitogenomic sequencing can be used for biodiversity assessment, presenting advantages over traditional approaches similar to those of metabarcoding, such as sample multiplexing. Additionally, mitogenomic sequencing provides increased information content through long mitochondrial contigs which contain multiple protein coding genes. These long contigs could provide improved phylogenetic resolution and measurement of intraspecific diversity at a more effective rate than single COI barcodes, while reducing the effects of false negatives caused by random drop out of genes due to degradation or variable sequencing coverage (Tang et al., 2014). The number of false positives could also be reduced by the use of longer sequences and selection on stringent criteria for establishing taxon presence. Furthermore, combinations of multiple markers increases delimitation success for closely related species compared to single marker work (Dupuis et al., 2012). Nevertheless, since multiple markers derived from mitochondrial reads represent a single linkage group, or could represent mito-nuclear discordance as a result of introgression (Weigand et al., 2017), the use of independent nuclear markers should also be

considered to resolve the presence of cryptic species (Campos-Soto, Torres-Pérez, & Solari, 2015; Miyamoto, Allard, Adkins, Janecek, & Rodney, 1994).

Metabarcoding-based biodiversity analysis

Metabarcoding has been mainly used for the recovery of species richness from community samples, uncovering in many cases extensive diversity which would have been difficult to achieve using traditional methods (Leray & Knowlton, 2015; Sinniger et al., 2016). Additionally, metabarcoding work is increasingly being proposed as an alternative to traditional ecosystem monitoring, where accurate estimations of species abundance in environmental samples are generally desirable (Ji et al., 2013; Shokralla et al., 2015). Sequencing read abundance, where higher proportion of species' biomass would be reflected by a higher proportion of sequencing reads has been suggested (e.g. Thomas, Deagle, Eveson, Harsch, & Trites, 2016), but these observations do not generally correspond to the majority of findings in the field (e.g. Elbrecht & Leese, 2015). Our results partially support previous findings, and mainly reflect the larger uncertainty of assumptions on relative species abundance as derived from metabarcoding workflows. More specifically, the metabarcoding work failed to detect significant relationships between read data and known biomass in the mock communities in several cases (Table 3). The FFFR amplicon data (universal Folmer primers) showed significant read-biomass relationships in only 4 out of 13 species, compared to 8 out of 13 for the other two amplicons. The discrepancy in efficiency between amplicons could be related to primer specificity or sequencing depth. Because the B1FR and FF130R primers were designed and modified for macroinvertebrate taxa, and secondly because the sequencing coverage achieved for the Folmer region (FFFR) compared to amplicon FF130R. Summing of sequencing results from all three amplicons slightly

improved the reads/biomass relationships, but mainly assisted in removal of false negatives from the metabarcoding data (Table 4).

Multi-dimensional scaling analysis (nMDS, Bray-Curtis index) of the metabarcoding data revealed similarities in community composition based on the sequencing results for individual amplicons (Figure S11 – S12). Such trends indicate that despite the variations in reads-biomass relationships for individual species, the community profiles obtained were still comparable, albeit with some exceptions (communities 9-10, Figure S11). When assessed against the shotgun data, similar patterns were found across treatments (individual amplicons), but the shotgun data were more condensed across the y-axis, resembling more closely results obtained from the B1FR amplicon (Figure S12).

Selection of the target region can influence metabarcoding results, and the utility of the COI marker has on occasion been questioned with regards to universality of the available priming sites (Deagle et al., 2014). Alternative COI metabarcoding primers have been designed for universal (Gibson et al., 2014; Hajibabaei et al., 2012; Leray et al., 2013), or group specific detection (e.g. freshwater invertebrates, Elbrecht & Leese, 2017; Gibson et al., 2014). Additionally, different markers have been proposed for use in characterisation of biodiversity through metabarcoding, such as 18S (Zhan et al., 2014), or 16S (Epp et al., 2012), though the COI still retains its superior value with some taxonomic groups compared to other markers due to the large repositories of reference sequences already available for a large number of taxa (Clarke et al., 2017). Increased accuracy in biodiversity detection in community samples could also be achieved through the simultaneous use of multiple amplicons. In Gibson *et al.* (2014), a set of 11 primer pairs targeting the COI barcoding region were used, showing that combinations of several primers significantly increased the levels of species detection in samples of known content. Our results partially support the

idea that the combination of sequencing reads from multiple amplicons can increase the detection rate of species richness of metabarcoding, as improvement of our results varied between the different species (Table 4). Similar findings are also reported by other multi-marker studies (Dupuis et al., 2012; Zhan et al., 2014). Furthermore, we should also take into consideration that the use of multiple amplicons or loci also creates additional costs for tagged primers and library preparation as well as handling and data analysis time (Creer et al., 2016).

False negatives, detection of rare diversity, and closely related species

The percentage of false negative detections for the shotgun work reached 5.6%, while false negative detections for metabarcoding were either comparable or somewhat lower per amplicon at 2.2% (B1FR), 4.5% (FF130R) and 5.6% (FFFR). However, metabarcoding was more unpredictable due to false negatives also occurring for species with higher biomass in the communities (e.g. 10 specimens of *E. danica*, FFFR amplicon). The shotgun method only missed rare species at the lowest end of biomass presence, which could be indicative of a need for higher sequencing depth for detection of rare species, as well as small bodied species with limited contribution to overall biomass. Primer binding related bias could have caused false negatives or abnormal biomass representation in metabarcoding data, through primer incompatibility (as was likely the case for species *E. danica*) or low sequencing depth, while variation in sequencing depth could also influence the quantitative relationships between reads and species biomass (Hajibabaei et al., 2011). Increased sequencing depth or use of multiple primers has been suggested in order to enhance the detection of species of smaller biomass or lower relative abundance in the samples through metabarcoding (Hajibabaei et al., 2012). Generally, the inability of either method to detect rare or low

biomass species could have significant implications for conservation surveys, as is the case for many endangered or invasive species (Zhan & MacIsaac, 2015).

Finally, the ability to discriminate between closely related species in community samples should also be taken into consideration. We analysed two closely related congener species of gastropods, *Bithynia leachii* and *Bithynia tentaculata*. Shotgun sequencing was able to effectively differentiate between the two species. Annotation of the reads for the two species by mapping onto the previously generated reference mitogenomes provided more confidence in the results. Similarly, identification of metabarcoding reads for the two species was achieved by BLAST identification of OTUs against individual barcode reference sequences. Using phylogenetic (NJ) analysis further supported the correct annotation of the sequences.

Future perspectives

Overall, the mitogenomic approach could present more effective and accurate detection of biomass and shifts in biomass in mixed bulk samples taken from the wild, compared to the more widely used to date COI metabarcoding. Nevertheless, the cost of metabarcoding currently remains lower than shotgun sequencing (in both cases multiplexing is reducing the analytical cost), and would allow metabarcoding to be used for applied ecosystem monitoring, especially in cases where accurate biomass information is not required, and good reference databases of the target groups exist. Additionally, current metabarcoding costs are comparable to morphology based identification, whilst providing comparable assessment results (e.g. stream bio assessment Elbrecht, Vamos, Meissner, Aroviita, & Leese, (2017)).

The importance of optimising and standardising such sequencing approaches is linked to generating more informative estimates of ecological interactions across taxa and trophic levels, as well as ecosystem functioning (Darling et al., 2017). Traditional monitoring is applied on individual or a limited number of species, which may not necessarily capture subtle responses to ecological change. Simultaneous monitoring across trophic levels, especially where the nature and dynamics of trophic interactions as well as other biotic interactions linked to competition and predation, might reveal more meaningful ecological signals relevant to biomonitoring (Woodward, Gray, & Baird, 2013). High throughput sequencing (HTS) data of species richness and relative abundance could be enhanced in the future through automated pipelines utilizing automated samplers and machine learning methods to reconstruct ecological networks and investigate ecological interactions at an unprecedented scale (Bohan et al., 2017). Such advances in technologies underpin the utility of HTS applications, both in generating comparable data sets across large spatial and temporal scales, as well as capturing increasingly subtle, but ecologically informative, signals of response to environmental change.

Acknowledgements

This work was funded and supported by the Environment Agency (EA) UK, and a Knowledge Economy Skills Scholarship (KESS). We also thank Bangor University for support and Wendy Grail for assistance and Andrian Chalkley for collection of invertebrate specimens used in experiments. Knowledge Economy Skills Scholarships (KESS) is a pan-Wales higher-level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West

Wales and the Valleys. We thank HPC Wales for allowing use of their systems for analysis. We also thank Prof. Florian Leese and two anonymous reviewers for comments on the manuscript. X.Z. is also supported by the Chinese Universities Scientific Fund (2017QC114) through China Agricultural University.

References

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., … Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. https://doi.org/10.1186/gb-2011-12-2-r18

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8), 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x

Benke, A. C., Huryn, A. D., Smock, L. a, & Wallace, J. B. (1999). Length-mass relationships for freshwater macroinvertebrates in North America with particular reference to the southeastern United States. *Journal of the North American Benthological Society*, *18*(3), 308–343. https://doi.org/10.2307/1468447

Bensasson, D., Zhang, D., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution ' s misplaced witnesses. *TRENDS in Ecology & Evolution*, *16*(6), 314–321. https://doi.org/10.5061/dryad.d2f79

Berry, D., Mahfoudh, K. Ben, Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, *77*(21), 7846–7849. https://doi.org/10.1128/AEM.05220-11

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., … Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*, 14087. https://doi.org/10.1038/ncomms14087

Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2017.03.001

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Campos-Soto, R., Torres-Pérez, F., & Solari, A. (2015). Phylogenetic incongruence inferred with two mitochondrial genes in Mepraia spp. and Triatoma eratyrusiformis (Hemiptera, Reduviidae). *Genetics and Molecular Biology*, *38*(3), 390–395. https://doi.org/10.1590/S1415-475738320140301

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high- throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. https://doi.org/10.1038/nmeth0510-335

Choo, L. Q., Crampton-Platt, A., & Vogler, A. P. (2017). Shotgun mitogenomics across body size classes in a local assemblage of tropical Diptera: Phylogeny, species diversity and mitochondrial abundance spectrum. *Molecular Ecology*, *17*(8), 2045–2054. https://doi.org/10.1111/mec.14258

Clarke, K. R., & Gorley, R. N. (2006). Primer v6: User Manual/Tutorial. Plymouth: PRIMER-E.

Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, *7*(3), 873–883. https://doi.org/10.1002/ece3.2667

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, pp. 1160–1170. https://doi.org/10.1111/1755-0998.12265

Collins, A., Ohandja, D. G., Hoare, D., & Voulvoulis, N. (2012). Implementing the Water Framework Directive: A transition from established monitoring networks in England and Wales. *Environmental Science and Policy*, *17*, 49–61. https://doi.org/10.1016/j.envsci.2011.11.003

Crampton-Platt, A., Timmermans, M. J. T. N., Gimmel, M. L., Kutty, S. N., Cockerill, T. D., Khen, C. V., & Vogler, A. P. (2015). Soup to tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a bornean rainforest sample. *Molecular Biology and Evolution*, *32*(9), 2302–2316. https://doi.org/10.1093/molbev/msv111

Crampton-Platt, A., Yu, D. W., Zhou, X., & Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, *5*(1), 15. https://doi.org/10.1186/s13742-016-0120-y

Cranston, P. S. (1990). Biomonitoring and invertebrate taxonomy. *Environmental Monitoring and Assessment*, *14*(2–3), 265–273. https://doi.org/10.1007/BF00677921

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., … Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, *56*, 68–74. https://doi.org/10.1111/2041-210X.12574

Darling, J. A., Galil, B. S., Carvalho, G. R., Rius, M., Viard, F., & Piraino, S. (2017). Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy*, *85*(May), 54–64. https://doi.org/10.1016/j.marpol.2017.08.014

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., Taberlet, P., Taberlet, P., …

Hajibabaei, M. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, *10*(9), 1789–1793. https://doi.org/10.1098/rsbl.2014.0562

Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., & Minamoto, T. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, *62*(1), 30–39. https://doi.org/10.1111/fwb.12846

Dupuis, J. R., Roe, A. D., & Sperling, F. A. H. (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology*, *21*(18), 4422–4436. https://doi.org/10.1111/j.1365-294X.2012.05642.x

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Elbrecht, V., & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE*, *10*(7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., & Leese, F. (2017). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, *5*(April), 1–11. https://doi.org/10.3389/fenvs.2017.00011

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.12789

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., … Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. *Molecular Ecology*, *21*(8), 1821–1833. https://doi.org/10.1111/j.1365-294X.2012.05537.x

Fahner, N. A., Shokralla, S., Baird, D. J., & Hajibabaei, M. (2016). Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. *PLoS ONE*, *11*(6), e0157505. https://doi.org/10.1371/journal.pone.0157505

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299. https://doi.org/10.1371/journal.pone.0013102

Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PLoS ONE*, *10*(10), 1–15.

https://doi.org/10.1371/journal.pone.0138432

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., … Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, *111*(22), 8007–8012. https://doi.org/10.1073/pnas.1406468111

Gillett, C. P. D. T., Crampton-Platt, A., Timmermans, M. J. T. N., Jordal, B. H., Emerson, B. C., & Vogler, A. P. (2014). Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, *31*(8), 2223–2237. https://doi.org/10.1093/molbev/msu154

Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M. J. T. N., Baselga, A., & Vogler, A. P. (2015). Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution*, *6*(8), 883–894. https://doi.org/10.1111/2041-210X.12376

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, *6*(4), e17497. https://doi.org/10.1371/journal.pone.0017497

Hajibabaei, M., Singer, G. a C., Clare, E. L., & Hebert, P. D. N. (2007). Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*, *5*(1), 24. https://doi.org/10.1186/1741-7007-5-24

Hajibabaei, M., Spall, J. L., Shokralla, S., & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, *12*(1), 28. https://doi.org/10.1186/1472-6785-12-28

Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*, S96–S99. https://doi.org/10.1098/rsbl.2003.0025

Hiiesalu, I., Pärtel, M., Davison, J., Gerhold, P., Metsis, M., Moora, M., … Wilson, S. D. (2014). Species richness of arbuscular mycorrhizal fungi: Associations with grassland plant richness and biomass. *New Phytologist*, *203*(1), 233–244. https://doi.org/10.1111/nph.12765

Hu, S. (1987). *Akaike information criterion statistics*. *Mathematics and Computers in Simulation* (Vol. 29). KTK Scientific Publishers, Tokyo. https://doi.org/10.1016/0378-4754(87)90094-2

Jackson, J. K., Battle, J. M., White, B. P., Pilgrim, E. M., Stein, E. D., Miller, P. E., & Sweeney, B. W. (2014). Cryptic biodiversity in streams: a comparison of macroinvertebrate communities based on morphological and DNA barcode identifications. *Freshwater Science*, *33*(1), 312–324. https://doi.org/10.1086/675225

Jeffery, N. W., & Gregory, T. R. (2014). Genome size estimates for crustaceans using Feulgen image analysis densitometry of ethanol-preserved tissues. *Cytometry Part A*, *85*(10), 862–868. https://doi.org/10.1002/cyto.a.22516

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Kelly, R. P., Port, J. a., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., … Crowder, L. B. (2014). Harnessing DNA to improve environmental management. *Science*, *344*(6191), 1455–1456. https://doi.org/10.1126/science.1251156

Krebes, L., & Bastrop, R. (2012). The mitogenome of Gammarus duebeni (Crustacea Amphipoda): A new gene order and non-neutral sequence evolution of tandem repeats in the control region. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *7*(2), 201–211. https://doi.org/10.1016/j.cbd.2012.02.004

Lacoursière-Roussel, A., Rosabal, M., & Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources*, *16*(6), 1401–1414. https://doi.org/10.1111/1755-0998.12522

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., … Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34. https://doi.org/10.1186/1742-9994-10-34

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., … Zhou, X. (2016). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, *16*(2), 470–479. https://doi.org/10.1111/1755-0998.12472

Loreau, M., & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: a synthesis of underlying mechanisms. *Ecology Letters*, *16*, 106–115. https://doi.org/10.1111/ele.12073

Macher, J.-N., Zizka, V. M. A., Weigand, A. M., & Leese, F. (2017). A simple centrifugation protocol for metagenomic studies increases mitochondrial DNA yield by two orders of

magnitude. *Methods in Ecology and Evolution*, *2017*(0), 1–5. https://doi.org/10.1111/2041-210X.12937

Mächler, E., Deiner, K., Steinmann, P., & Altermatt, F. (2014). Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science*, *33*, 1174–1183. https://doi.org/10.1086/678128.

Mährlein, M., Pätzig, M., Brauns, M., & Dolman, A. M. (2016). Length–mass relationships for lake macroinvertebrates corrected for back-transformation and preservation effects. *Hydrobiologia*, *768*(1), 37–50. https://doi.org/10.1007/s10750-015-2526-4

Meusnier, I., Singer, G. A. C., Landry, J. F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, *9*(1), 214. https://doi.org/10.1186/1471-2164-9-214

Miyamoto, M. M., Allard, M. W., Adkins, R. M., Janecek, L. L., & Rodney, L. (1994). A Congruence Test of Reliability Using Linked Mitochondrial DNA Sequences. *Systematic Biology*, *43*(2), 236–249. https://doi.org/10.1093/sysbio/43.2.236

Morinière, J., Hendrich, L., Balke, M., Beermann, A. J., König, T., Hess, M., … Haszprunar, G. (2017). A DNA barcode library for Germany's mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Molecular Ecology Resources*, *17*(6), 1293–1307. https://doi.org/10.1111/1755-0998.12683

O'donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template- Specific bias in Large-Scale DNA sequencing studies. *PLoS ONE*, *11*(3), 1–11. https://doi.org/10.1371/journal.pone.0148698

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Pfrender, M., Hawkins, C., Bagley, M., Courtney, G., Creutzburg, B., Epler, J., … Whiting, M. (2010). Assessing Macroinvertebrate Biodiversity in Freshwater Ecosystems: Advances and Challenges in DNA-based Approaches The Quarterly Review of Biology. *Source: The Quarterly Review of Biology*, *85*(3), 319–340. https://doi.org/10.1086/655118

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*(4), 819–830. https://doi.org/10.1111/1755-0998.12355

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–25. https://doi.org/citeulike-article-id:93683

Seymour, M., Durance, I., Cosby, B. J., Ransom-Jones, E., Deiner, K., Ormerod, S. J., … Creer, S. (2018). Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, *1*(1), 4. https://doi.org/10.1038/s42003-017-0005-3

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., … Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports*, *5*, 9687. https://doi.org/10.1038/srep09687

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A. J., Yamamoto, H., Chevaldonné, P., … Creer, S. (2016). Worldwide Analysis of Sedimentary DNA Reveals Major Gaps in Taxonomic Knowledge of Deep-Sea Benthos. *Frontiers in Marine Science*, *3*, 92. https://doi.org/10.3389/fmars.2016.00092

Sweeney, B. W., Battle, J. M., Jackson, J. K., & Dapkey, T. (2011). Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? *Journal of the North American Benthological Society*, *30*(1), 195–216. https://doi.org/10.1899/10-016.1

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596–1599. https://doi.org/10.1093/molbev/msm092

Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., … Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, *6*(9), 1034–1043. https://doi.org/10.1111/2041-210X.12416

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., … Zhou, X. (2014). Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, *42*(22), e166–e166. https://doi.org/10.1093/nar/gku917

Team, R. C. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. R Foundation for Statistical Computing.

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2016). Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, *16*(3), 714–726. https://doi.org/10.1111/1755-0998.12490

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., & Leese, F. (2017). Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology*, *26*(20), 5705–5715. https://doi.org/10.1111/mec.14292

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., … Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 27. https://doi.org/10.1186/s40168-017-0237-y

Woodward, G., Gray, C., & Baird, D. J. (2013). Biomonitoring for the 21st Century: new perspectives in an age of globalisation and emerging environmental threats. *Limnetica*, *32*(2), 159–174.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., … Wang, J. (2014). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Yang, C., Wang, X., Miller, J. A., De Blécourt, M., Ji, Y., Yang, C., … Yu, D. W. (2014). Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. *Ecological Indicators*, *46*, 379–389. https://doi.org/10.1016/j.ecolind.2014.06.028

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zhan, A., Bailey, S. A., Heath, D. D., & Macisaac, H. J. (2014). Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources*, *14*(5), 1049–1059. https://doi.org/10.1111/1755-0998.12254

Zhan, A., & MacIsaac, H. J. (2015). Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. *Conservation Genetics*, *16*(3), 513–522. https://doi.org/10.1007/s10592-014-0678-9

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., … Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, *2*(1), 4. https://doi.org/10.1186/2047-217X-2-4

Data accessibility

The data for this work can be accessed at the European Nucleotide Archive (ENA) database (Study accession PRJEB23036).

Author contributions

IB, SC, GRC designed the experiments. KW, MH, XZ, MC contributed to experimental design. IB performed morphometric measurements, optimised and performed DNA extractions and metabarcoding library prep, ran bioinformatics of metabarcoding data. MS and IB performed statistical analysis. SS, MH ran the metabarcoding sequencing. MT, SL, XZ ran

sequencing and bioinformatics analysis of mitogenomics. DB taxonomically identified invertebrate specimens. IB wrote first version of manuscript. SC, GRC edited manuscript, and all authors provided comments.

Table 1: Design of mock communities. See columns for the detailed contents of each community (1-10), with numbers referring to specimens from each species used. Last column: total number of specimens/species, bottom line: total number of specimens/community and number of species/community. Highlighted the species with lowest abundance (dark grey) and highest abundance (light grey) in each community. *Drosophila melanogaster* was used as positive control for each mock community.

		Community										
Number	**Species**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Specimens per species**
1	*Anisus vortex*	35	40	45	5	25	20	15	10	30	1	226
2	*Asellus aquaticus*	1	4	8	10	14	17	19	21	24	24	142
3	*Bathyomphalus contortus*	14	13	12	11	10	8	6	1	2	4	81
4	*Bithynia tentaculata*	24	10	6	25	26	1	27	15	20	26	180
5	*Ephemera danica*	16	3	1	6	8	12	10	18	14	20	108
6	*Gyrinus marinus*	2	1	3	10	4	8	5	9	6	7	55
7	*Planorbis planorbis*	24	25	19	22	1	4	7	10	13	16	141
8	*Potamopyrgus antipodarum*	10	32	28	25	21	33	14	17	1	5	186
9	*Radix balthica*	3	15	5	17	16	10	12	1	9	6	94
10	*Physa fontinalis*	1	3	4	6	8	10	12	13	13	0	70
11	*Notonecta glauca*	10	0	0	4	2	1	0	6	0	8	31
12	*Bithynia leachi*	12	3	5	1	9	11	8	7	13	14	83
13	*Gammarus pulex*	2	5	6	4	8	8	1	8	3	7	52
14	*Drosophila melanogaster*	3	3	3	3	3	3	3	3	3	3	30
	Total specimens	**157**	**157**	**145**	**149**	**155**	**146**	**139**	**139**	**151**	**141**	**1479**
	Total Number of species	**14**	**13**	**13**	**14**	**14**	**14**	**13**	**14**	**13**	**13**	

Table 2: Estimated biomass for each species included in the mock communities. Values are presented in milligrams (mg).

	Community									
Species	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Anisus vortex	5.08	5.75	6.44	0.70	3.41	2.84	2.10	1.42	4.21	0.13
Bathyomphalus contortus	0.79	0.70	0.65	0.56	0.47	0.40	0.31	0.05	0.10	0.22
Planorbis planorbis	8.87	9.04	6.75	7.66	0.37	1.56	3.10	4.53	4.77	5.55
Bithynia leachi	11.97	3.19	4.52	0.94	8.95	10.30	7.45	6.61	12.50	14.54
Bithynia tentaculata	115.60	46.33	26.58	113.80	128.77	4.01	135.45	79.47	104.71	124.64
Physa fontinalis	1.16	4.14	4.78	7.91	10.81	13.10	16.42	17.02	18.70	**0.00**
Potamopyrgus antipodarum	3.82	13.37	11.66	9.89	8.10	13.87	5.27	6.46	0.36	1.74
Radix balthica	20.47	89.72	31.37	106.12	101.69	57.27	70.67	6.08	53.80	35.91
Notonecta glauca	77.38	**0.00**	**0.00**	33.88	16.09	7.47	**0.00**	46.54	**0.00**	62.09
Asellus aquaticus	1.00	4.92	9.00	13.02	19.50	22.01	22.68	27.56	30.73	37.72
Gammarus pulex	30.49	54.70	75.11	51.18	90.38	104.72	10.13	128.80	42.39	83.34
Ephemera danica	71.92	14.24	4.18	28.62	35.93	54.17	46.02	80.35	62.88	86.31
Gyrinus marinus	21.69	9.36	30.00	104.94	40.35	85.77	52.20	98.29	60.07	72.01
Total (mg)	**370.25**	**255.46**	**211.02**	**479.23**	**464.82**	**377.48**	**371.80**	**503.18**	**395.20**	**524.22**

Table 3: COI primers used for metabarcoding.

Primer Name	Primer Sequence	Direction	Citation
LCO1490	GGTCAACAAATCATAAAGATATTGG	F	Folmer *et al.* 1994
HC02198	TAAACTTCAGGGTGACCAAAAAATCA	R	Folmer *et al.* 1994
I-B1	CCHGATATAACITTYCCICG	F	Hajibabaei *et al* . 2012 (modified)
I-130R	GAAAATYATAAIGAAIGCRTGAGC	R	Meusnier *et al.* 2008 (modified)

Table 4: Summary table of significance of reads to biomass correlations, for each sequencing treatment. Amplicon data ("Amplicon", B1FR:450bp, FF130R: 130bp, FFFR: 658bp, SumAmplicon: sum of all amplicon data per species), and shotgun data ("Shotgun", pShotgun: proportion of reads, MitoNorm: mito-ratio normalised). The type of model used is indicated by shading (light grey: logistic, white: linear). Shotgun data not shown for species *A. vortex* (NA). Significant relationships (<0.05) are indicated with (*).

Number	Taxa		Amplicon				Shotgun	
	Family	Species	B1FR	FF130R	FFFR	SumAmplicon	pShotgun	MitoNorm
1	Planorbidae	*Anisus vortex*	0.01*	<0.01*	0.06	0.02*	NA	NA
2	Planorbidae	*Bathyomphalus contortus*	<0.01*	0.06	0.07	0.02*	<0.01*	<0.01*
3	Planorbidae	*Planorbis planorbis*	0.03*	0.06	0.09	0.07	0.01*	0.01*
4	Bithyniidae	*Bithynia leachi*	0.11	0.01*	0.04*	0.03*	<0.01*	<0.01*
5	Bithyniidae	*Bithynia tentaculata*	0.58	0.27	0.37	0.46	0.08	0.09
6	Physidae	*Physa fontinalis*	0.25	<0.01*	0.57	0.04*	<0.01*	<0.01*
7	Hydrobiidae	*Potamopyrgus antipodarum*	0.04*	<0.01*	0.02*	0.02*	<0.01*	<0.01*
8	Lymnaeidae	*Radix balthica*	0.03*	0.01*	0.03*	0.01*	<0.01*	0.01*
9	Notonectidae	*Notonecta glauca*	0.02*	0.01*	0.51	0.09	0.01*	0.02*
10	Asellidae	*Asellus aquaticus*	0.05*	0.07	0.15	0.06	0.02*	0.03*
11	Gammaridae	*Gammarus pulex*	0.52	0.06	0.32	0.46	<0.01*	<0.01*
12	Ephemeridae	*Ephemera danica*	0.04	0.02*	0.06	0.02*	<0.01*	<0.01*
13	Gyrinidae	*Gyrinus marinus*	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	0.01*